AD_____

Award Number: DAMD17-01-1-0516

TITLE: A Constraint Satisfaction Neural Network Approach for Data Mining Classification and Association Rules in Breast Cancer Databases

PRINCIPAL INVESTIGATOR: Georgia D. Tourassi, Ph.D.

CONTRACTING ORGANIZATION: Duke University Medical Center
Durham, North Carolina 27710

REPORT DATE: January 2003

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	January 2003	Final (1 Sep 01 - 31 Dec 02)

4. TITLE AND SUBTITLE
A Constraint Satisfaction Neural Network Approach for Data Mining Classification and Association Rules in Breast Cancer Databases

5. FUNDING NUMBERS
DAMD17-01-1-0516

6. AUTHOR(S) :
Georgia D. Tourassi, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Duke University Medical Center
Durham, North Carolina 27710

E-Mail: gt@deckard.mc.duke.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. Abstract ***(Maximum 200 Words)*** *(abstract should contain no proprietary or confidential information)*

We propose to explore an innovative, data mining (DM) process for application in breast cancer (BC) databases. The DM process is the Constraint Satisfaction Neural Network (CSNN). Contrary to feed-forward networks and statistical models, the CSNN has a non-hierarchical architecture that allows it to be used either as a prediction/classification tool or as an analysis tool for mining association rules in databases. This is a feasibility study to investigate to what degree the CSNN can deliver the above promises for the mammographic diagnosis of breast lesion malignancy.

The main objectives of the study are the following: (1) to develop a CSNN for mining a database of patients suspected with BC who underwent breast biopsy; (2) to evaluate the CSNN as a diagnostic tool; (3) to evaluate the CSNN as a patient prototype analysis tool to discover prevalent trends and associations among the variables; (4) to assess the network's robustness with missing data.

Initially, the CSNN is intended as a computer-assisted diagnostic tool to help physicians optimize the decision to refer a probably benign breast lesion to short-term follow-up instead of biopsy. Ultimately, the CSNN can be applied as a support tool individualizing the decision process in BC patient management.

14. SUBJECT TERMS:	15. NUMBER OF PAGES
constraint satisfaction, computer-aided diagnosis, neural networks, breast cancer	52
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

1. Introduction

The purpose of this study was to explore an innovative data mining (DM) process for application in a breast cancer database. The DM process is the Constraint Satisfaction Neural Network (CSNN). CSNNs are typically used for solving system optimization problems [1-4]. The proposed study is based on the hypothesis that mammographic diagnosis can be approached as a system optimization problem. Accordingly, a patient is modeled as a non-linear, dynamic system comprised of several components (e.g. clinical findings, personal and family history, mammographic findings, presence or absence of breast cancer). All components are coded into variables interconnected with constraints to keep the system stable. Typically, there is information about some system components (e.g. clinical and mammographic findings) and some questions need to be answered (e.g. Is there breast cancer?). Answering such questions is equivalent to finding the optimal values for the corresponding variables (i.e., lesion malignancy) so that the constraints are satisfied to a maximum extent and the system is stable. The CSNN is designed to solve such optimization problems. Furthermore, CSNN's non-hierarchical architecture allows it to be used not only as a prediction tool but also as an analysis tool for patient profiling. The proposed study aims to explore both promises with a breast cancer database.

2. Body

Statement of Work

This is the final report for this project, which was originally a one-year concept project scheduled for completion by August 31, 2002. During the funded period, a no-cost extension (till December 31, 2002) was approved to accomplish the following specific aims:

1. Develop a CSNN for mining a database of patients suspected with breast cancer who underwent breast biopsy;
2. Evaluate the CSNN as a diagnostic tool;
3. Evaluate the CSNN as a patient prototype analysis tool to discover prevalent trends and associations among the variables;
4. Assess the network's robustness with missing data.

The accomplishments of the entire effort will be summarized based upon these aims

Overview of Progress for Each Aim

Aim 1. First, we developed a CSNN to predict breast cancer malignancy based on patient's clinical findings and the breast lesion's mammographic presentation.

Database: We utilized a database of consecutive patients who presented to diagnostic mammography with non-palpable breast lesions and referred for biopsy (core or excisional) at Duke Medical Center from 1991 to 2000. There were in total 1,530 breast lesions with definitive histopathological diagnosis.. The mean patient age was 56 years, with a range of 23-89 years. The database included 1,530 breast lesions of which 533 (35%) were found to be malignant. The database contained 715 cases with masses, including 83 cases with calcifications in addition to a mass. There were 674 cases with calcifications only. The malignancy rate for masses and calcifications is similar (36% and 34% respectively). There were 141 cases with neither a mass nor a calcification. These lesions were reported as special findings (architectural distortion, focal asymmetric density, etc.).

Each breast lesion was described using ten mammographic findings from the BI-RADS™ lexicon [5]. The findings were: mass size, mass margin, mass density, mass shape, calcification description, calcification number, calcification distribution, special cases, associated findings, and quadrant location of abnormality. Six findings from the patient's medical history (age, menopausal status, use of replacement hormones, previous history of breast cancer, previous benign biopsy for breast cancer, family history of breast cancer) were also included in the database. In addition, for each case the database contained the attending mammographer's assessment for the likelihood of malignancy on a scale of 1 to 5. Note that this is different in form and intent from the BI-RADS™ clinical assessment. Finally, the malignant/benign result for each lesion was abstracted from the pathology report and was entered into the research database. Complete mammographic and clinical findings were available for all 744 breast lesions in the dataset. For the remaining 786 lesions, there were only mammographic findings

available plus the patient's age at the time of diagnosis. The remaining clinical and history findings were unavailable for those patients.

Data preprocessing: We converted the mammographic and clinical findings into a binary input vector. For each patient, the input vector consisted of exclusively binary nodes 0 or 1 representing if a particular finding is present or not. The input findings were coded so that one neuron is assigned to each possible description for every finding. For example, according to the BI-RADS™ lexicon, there are four possible types of "mass shape". Four nodes were assigned to this BI-RADS™ finding, each one corresponding to a different type of mass shape (i.e., round, oval, lobulated, irregular). In addition, four separate neurons were assigned to correspond to the presence of masses, microcalcifications, special findings, and associated findings. The continuous findings such as patient age and mass size were represented as categorical data as well by bining the continuous values. Finally, one extra node was added to constitute the diagnosis. The diagnosis neuron took the value of 1 if breast cancer is present and the value of 0 if breast cancer is absent. We used only one neuron to represent diagnosis so that the CSNN can be used as a predictive rather than a classification tool. Binary format makes the data mining process easier.

CSNN architecture: The CSNN is a Hopfield-type network [6]. The network consists of neurons arranged in a non-hierarchical structure. Therefore, contrary to traditional predictive models, the CSNN does not have designated input and output neurons. The neurons are highly interconnected with symmetrical, bidirectional weights ($w_{ij}=w_{ji}$). Given an optimization problem, the CSNN weights w_{ij} can be interpreted as the problem constraints and every network state can be viewed as a possible solution. The CSNN network operates as a non-linear, dynamic system aimed to achieve global stability by assigning values to its neurons while the weights remain fixed. To achieve global stability, the CSNN employs a dynamic and iterative mechanism. The mechanism assumes that the activation level of all neurons can take any value in the range [0,1]. The CSNN is designed to maximize the activation of its neurons in relation to the constraints existing among them. To achieve this goal, the activation level of each neuron i is updated using the delta rule introduced by Rumelhart [7]. With this update rule, the network will restrict the activation levels to the [0,1] range and will evolve so that all neurons achieve their maximum possible activation while still satisfying the constraints imposed by the weights. The measure of global stability is a Lyaponov function E (referred to as "energy function") often used to describe the state of nonlinear dynamic systems [6]:

- $$E_{(n)} = -\frac{1}{2}\sum_i\sum_j w_{ij}\cdot u_{i(n)}\cdot u_{j(n)} - \sum_i Bias_i\, u_{i(n)} + \sum_i Ext_i\cdot u_{i(n)}.$$

A dynamic system achieves a stable state when this function is minimized. In the CSNN context, the energy function is a measure of constraint satisfaction. The first two terms of E describe the internal dynamics of the network. The last term is the penalty imposed by the external influences.

A crucial step for developing a CSNN is determining the constraints weight matrix. The weight matrix contains the relations or constraints among all neurons. For this study we explored an autoassociative backpropagation (auto-BP) scheme that showed great potential before [8,9] and was also utilized in our pilot study [10]. When the training phase is complete, the autoassociative BP weights act as the CSNN constraints

satisfying the main conditions. Utilizing a backpropagation scheme to determine the CSNN constraints is highly innovative, overcoming the limitations of hard constraints typically associated with constraint satisfaction problems [1].

CSNN implementation: To accomplish specific aim 1, we implemented a CSNN with a total of 83 neurons. Each neuron represented a different description for every mammographic and clinical finding included in the database (as described in the data preprocessing section). One neuron was assigned to describe the malignancy status of a lesion. An auto-associative backpropagation neural network was also implemented to determine the CSNN constraints (as described in the previous section).

Aim 2. Then, we applied the CSNN as a diagnostic tool for prediction of the breast biopsy outcome. In addition, we studied the effect of data sampling in the overall diagnostic performance of the CSNN.

(1) First, we applied a 50%-50% cross-validation sampling scheme. The dataset was randomly divided in two subsets (A and B). Initially, subset A was used to determine the CSNN constraints by applying the auto-associative backpropagation scheme described before. Then, the predictive ability of the CSNN was tested on subset B. For each test case, CSNN proceeded iteratively until its energy function was stabilized. At the end of the iterative process, the activation level achieved by the designated diagnosis neuron was used as a decision variable for Receiver Operating Characteristics (ROC) analysis. Then, the whole process was reversed so that subset B was used to determine the CSNN constraints and subset A was used to test the CSNN as a predictive tool. We used the ROCKIT software package developed by Metz *et al.* [11] to fit ROC curves to the final activation level of the CSNN diagnosis neuron. Table 1 compares the CSNN to experienced mammographers. Several indices of diagnostic performance are presented: overall ROC area index, specificity at 95% sensitivity level, and the corresponding PPV at the same operating point. The ROC evaluation of the mammographers' performance was based on a gestalt, 5-point scale, categorical assessment of the likelihood of malignancy (NOT the BI-RADS assessment).

Table 1

CAD MODEL	ROC Area Index	SPECIFICITY	PPV
CSNN	0.83 ± 0.03	47%	49%
Mammographers	0.82 ± 0.02	37%	45%

(2) Since our dataset spans almost a decade, we studied if there any differences in CSNN performance due to changes in patient population at our institution. First, we trained the CSNN on the initial 500 lesion (biopsied between 1991 and 1996). Then, we tested if the CSNN can achieve clinically acceptable diagnostic accuracy on the remaining 1,030 cases (biopsied between 1996 and 2000). It needs to be emphasized that from the 1,030 test cases, only 244 had complete mammographic and clinical findings. The remaining 786 cases had missing data.

The results of the validation study are summarized in Table 2. The table shows the overall ROC area index A_z of the CSNN along with the partial area above a sensitivity

of 90% ($_{0.90}A_Z$). The partial ROC area index for the high sensitivity range is a clinically more meaningful performance index for this diagnostic problem. The table also includes the CSNN specificity at 95% sensitivity. For comparison the Table also includes the previously reported CSNN performance on the initial 500 cases according to a 50%-50% cross-validation sampling scheme.

Table 2: Diagnostic Performance of the CSNN on the Train and Validation Sets

Data Set	A_Z ±STD	$_{0.90}A_Z$±STD	PPV at 95% Sensitivity
Initial (500 cases)	0.84±0.02	0.35±0.06	50%
Validation (1,030 cases)	0.81±0.02	0.26±0.03	41%

In addition, the CSNN performance was analyzed separately according to the types of breast lesions (Table 3). Previous studies with a variety of artificial intelligence techniques have demonstrated that diagnostic performance varies substantially between masses and classifications. Specifically, CAD performance on breast masses is superior to that on calcifications. Similar trend was observed in our validation study as well. CSNN performance was significantly better on masses than on calcifications. However, compared to results on the initial 500 cases, the CSNN performance deteriorated slightly on masses but improved on calcifications.

Table 3: CSNN diagnostic performance based on the type of lesions present.

Type of Lesions	No. of Cases in Initial Set (% maligancy)	No. of Cases in Validation Set (% maligancy)	A_Z INITIAL SET	A_Z VALIDATION SET
Masses only	232 (29.7%)	402 (35.6%)	0.93±0.02	0.90±0.02
Calcifications only	192 (37.5%)	483 (31.5%)	0.65±0.04	0.70±0.03
Masses with Calcifications	29 (62.1%)	54 (50.0%)	0.83±0.08	0.75±0.07
No Masses or Calcifications	47 (31.9%)	91 (38.5%)	0.70±0.09	0.82±0.05

The above validation study was accepted for oral presentation at the 2003 SPIE Meeting in Medical Imaging, San Diego CA, February 15-20. A copy of the conference proceedings article is provided in the appendix (item A).

Aim 3. Next, we evaluated the CSNN as a patient prototype analysis tool to discover prevalent trends and associations among the variables.

This is the most exciting aspect of the study. The ability to use the network from "bottom-up" is very attractive compared to the backpropagation neural network. By selecting the neurons that accept external information, the CAD tool operator can interrogate the constraint satisfaction network and get more detailed explanations of its decision reasoning. We demonstrated this quality by asking the network three prototypical questions. Given our database (i.e. lesions sufficiently suspicious of breast cancer to require mammographers recommend biopsy), what is the profile of a patient with breast cancer (BC)? What is the profile of a patient without BC? What is the profile of a "confusing" patient? Table 2 summarizes the results. For each prototype, only the activated variables are displayed.

Table 4

Activated Variables	BC PRESENT	BC ABSENT	UNCERTAIN
Masses	Yes		
margin	Spiculated		
shape	Irregular		
density	High		
size	<10mm		
Associated Findings	architectural distortion		
Age	>70 yrs	40-50 yrs	>70 yrs
Menopausal Status	Post	Post	Post
Family Hx of BC	Yes		Yes

To answer the first question ("What is the profile of a patient with BC?"), the activation level of the diagnosis neuron was set to 1.0 and the remaining neurons were left free to evolve until the network reached a stable state. None of the remaining neurons accepted external input. This is equivalent to using a backpropagation neural network from the "bottom-up". Table 2 shows which neurons were activated and reached maximum values indicating strong association with the particular diagnosis outcome (i.e., BC present, BC absent, Uncertain).

The following hidden associations were discovered:

1. The mammographic variables strongly correlated with cancer are architectural distortion and small, spiculated masses with irregular shape, and high density.
2. The profile of a patient without breast cancer, is a younger female (40-50 yrs old), menopausal, without any family or personal history of BC, and without any lesions. Although it is not clinically surprising that the patient without BC has no masses or calcifications present, it is unexpected given that the majority of the cancer free patients in the database had some type of lesion present. This is an indication that the CSNN gives responses that are not necessarily statistical in nature.

3. The profile of a "confusing" patient is also very informative. To acquire this profile, the activation level of the diagnosis neuron was set at 0.5. The neurons that were activated were the same ones as in the BC prototype with the exception of masses. Therefore, older age and family history of BC alone increase the risk of breast cancer, as it is clinically known.

The data mining process can go one step further by controlling externally more than one neuron at a time. Table 3 shows the results of this process for breast cancer patients according to their breast lesion type (mass vs. calcification). To acquire those patient prototypes, three neurons (BC present, mass absent, calcification absent) were externally controlled and the remaining 80 neurons evolved until the network reached a stable state. Specifically, the second column shows that clustered, pleomorphic calcifications, architectural distortion, and focal asymmetric density are strongly associated with breast cancer. In the absence of masses or calcifications, the presence of architectural distortion and/or focal asymmetric density are high risk factors for breast cancer.

Table 5

	BC PRESENT		
Activated Variables	**No Masses**	**No Masses, No Calcifications**	**No Masses, No Calcifications, No Focal Asymmetric Density**
Calcifications	Yes	No	No
distribution	clustered		
number	>10		
description	pleomorphic		
Associated Findings	architectural distortion	architectural distortion	architectural distortion
Special Findings	focal asymmetric density	focal asymmetric density	No
Family Hx of BC	Yes	Yes	Yes

The potential to use CSNN for data profiling is very exciting and we chose to explore it further. Specifically, we applied the CSNN to profile patient clusters. Initially, a self-organizing map (SOM) was used to identify clusters in a large, heterogeneous computer-aided diagnosis database based on mammographic findings (BI-RADS™) and patient age. The resulting clusters were then characterized by their prototypes determined using the CSNN. The patient clusters showed logical separation of clinical subtypes such as architectural distortions, masses, and calcifications. Moreover, the study showed such identification and profiling of subgroups within a database could help elucidate clinical trends and facilitate future decision model building. Specifically, the study showed that broad categories of masses and calcifications were stratified into several clusters. The percent of the cases that were malignant was notably different among the clusters. A feed-forward back-propagation artificial neural network (BP-

ANN) was used to identify likely benign lesions that may be candidates for follow up rather than biopsy. The performance of the BP-ANN varied considerably across the clusters identified by the SOM. In particular, a cluster was identified that accounted for 79% of the recommendations for follow up that would have been made by the BP-ANN.

The above study where the CSNN was utilized for patient profiling has been accepted for publication in *Artificial Intelligence in Medicine*. A copy of the manuscript is provided in the appendix (item B).

Aim 4. Finally, we assessed the CSNN's robustness with missing data.

As explained in the data description, the majority (786/1,030) of breast lesions in our database are missing the patients' clinical findings. The ROC area index of the CSNN was evaluated separately on the cases with complete findings and on those with incomplete clinical findings. As expected, the ROC area index was lower in cases with missing data (A_z=0.80±0.02) than in those with complete findings (A_z=0.84±0.03). The difference was statistically significant at the 95% confidence level. Similar trend was observed with the partial ROC area indices.

The next table presents the effect of missing data in detail, according to the type of breast lesions present. The table shows that the missing data did not affect the overall performance of the network on "mass only" cases. The CSNN performance on calcifications was slightly better for lesions with complete findings, however the difference was not statistically significant. A notable difference in performance was observed for the "masses with calcifications" category. However, the small number of cases in this category does not allow conclusive remarks. This is also the case for lesions without masses or calcifications present ("neither").

Table 6: CSNN performance according to the type of lesions present for cases with complete and incomplete findings

	Masses only	Calcifications only	Masses + Calcifications	Neither	ALL
Train	0.93 ± 0.02	0.65 ± 0.04	0.83 ± 0.08	0.70 ± 0.09	0.84 ± 0.02
Validation	0.88 ± 0.02	0.70 ± 0.03	0.75 ± 0.07	0.82 ± 0.05	0.81 ± 0.02
Complete	0.88 ± 0.03	0.73 ± 0.07	1.0	0.81 ± 0.12	0.84 ± 0.03
Incomplete	0.87 ± 0.02	0.70 ± 0.03	0.63 ± 0.09	0.83 ± 0.05	0.80 ± 0.02

The non-hierarchical architecture of the CSNN makes possible its utilization on cases with partially missing data. Other predictive models require an additional technique to impute the missing data before a case is tested. Contrary, the CSNN does not require

such step. Specifically, the CSNN can be applied to reconstruct simultaneously not only the correct diagnosis but also any missing components of a given clinical case. This is an exciting possibility for clinical databases with missing data. Imputing missing data is an important issue that tends to compromise the performance of a decision model.

We tested the accuracy of the CSNN to impute missing data while performing as a diagnostic tool. We focused on imputing the patient age. Previous studies have shown that the patient age is the strongest predictive clinical factor of malignancy . We tested the CSNN on the same 1,030 validation cases. However, the CSNN neurons that represent patient age were left to evolve without any external influences. Therefore, we simulated an experiment where the CSNN was asked to perform as a diagnostic tool while imputing simultaneously a very important predictive finding (i.e., patient age).

Although the overall performance of the CSNN deteriorated (A_z=0.78±0.02), it was still able to predict breast lesion malignancy with sufficient accuracy. Furthermore, the CSNN was able to impute the missing patient age accurately in 30% of the cases. In 69% of the cases, the CSNN imputed patient age within adjacent age groups. Table 6 summarizes the results of this experiment. The table shows the correct and CSNN predicted age groups for all patients in the validation set.

Table 7: CSNN accuracy on imputing the missing patient age while performing the diagnostic task

Patient Age Groups	No. of cases in each age group	Accuracy	Accuracy (± 1 age group)
<= 40 yrs	82	14.6%	51.2%
(40,50]	321	32.0%	69.5%
(50-60]	276	37.3%	68.8%
(60,70]	196	18.4%	85.7%
>70 yrs	155	34.8%	56.8%
TOTAL	1,030	29.9%	69.0%

The above results are included in the SPIE proceedings article that is provided in the appendix.

3. Key Research Accomplishments

This research resulted in the following major accomplishments:

- By modeling decision making as a system optimization problem, we were able to utilize an innovative neural network that views the patient as a dynamic system without designated input and output neurons (non-hierarchical architecture). The CSNN can predict the activation status of some neurons based on the known status of the remaining neurons and the known nature of interactions in the system. We demonstrated this quality by applying the CSNN as a decision support tool to decide the malignancy status of a suspicious breast lesion. The CSNN was very effective as a diagnostic tool showing performance similar to that of experienced mammographers.

- Its non-hierarchical architecture allowed the CSNN to be utilized as a knowledge discovery tool. Decoding hidden data trends and associations may help physicians understand better and refine established clinical judgment patterns. We demonstrated this potential by applying the CSNN as a patient profiling tool.

- We demonstrated the potential of applying the CSNN as a computer aid to improve upon the diagnostic accuracy of the radiologists for their decision to biopsy a suspicious breast lesion. While maintaining 95% sensitivity for cancers, the model could have obviated 47% of the benign biopsies.

- Finally, we evaluated the impact of missing data in the overall diagnostic performance of the CSNN. In addition, the CSNN ability to effectively impute missing clinical data while performing as a predictive tool was verified successfully.

4. Reportable Outcomes

Publications:

The following publications resulted directly from this work. They consist of a peer-reviewed journal article and a conference proceedings article. They have both been accepted for publication. Copies are attached as appendices A and B.

1. M.K Markey, J.Y. Lo, **G.D. Tourassi**, C.E. Floyd, Jr., "Self-Organizing Map for Cluster Analysis of a Breast Cancer Database," accepted for publication in *Artificial Intelligence in Medicine* [12/02].
2. **G.D. Tourassi**, M.K. Markey, and J.Y. Lo, "Validation of a constraint satisfaction neural network for breast cancer diagnosis: New results from 1030 cases," accepted for oral presentation at the 2003 SPIE Medical Imaging Conference, San Diego, CA, 15-20 February.

Personnel Receiving Salary:

1. Georgia D. Tourassi, PhD, (PI)

Funding:

The following applications for funding resulted directly from this research.

1. Idea Award, NIH/NCI, "Computer-assisted recommendation to breast biopsy," PI Georgia D. Tourassi, co-investigators Baker JA, Lo JY, et. al., direct costs $250,000, 4/1/03–3/31/05. The grant application received a non-fundable score (240).

5. Conclusions

This research resulted in several major advancements in the fields of data mining and breast imaging. At present, diagnostic mammograms are used to determine if a suspicious breast lesion should be sent to biopsy or follow-up. However, the majority of biopsies (65-85%) performed due to suspicious mammograms are found to be benign [12,13]. Several attempts have been made to develop computer-assisted diagnostic (CAD) models that help physicians improve the cost-effectiveness of their decision to send a suspicious breast lesion to biopsy [14,15]. Most CAD efforts focus on traditional backpropagation neural networks and expert systems.

In this project, we investigated a novel CAD approach that exploits non-linear dynamic system principles. The CAD technique is based on the Constraint Satisfaction Neural Network (CSNN) and it approaches breast cancer diagnosis as a constraint satisfaction problem. Our study demonstrated that the CSNN combines three important qualities for a successful CAD system:

(1) Accuracy - the CSNN performed as well as other CAD models previously published in the literature.

(2) Interpretability - Its non-hierarchical architecture allows the CSNN to be utilized as a knowledge discovery tool. Decoding hidden data trends and associations helps make the CSNN decision making process transparent to the user, thus facilitating clinical acceptance.

(3) Adaptability - For a given patient, the CAD operator can choose which neurons need prediction without further re-training or re-organizing of the neural network. Consequently, the CSNN is an appropriate decision model for clinical databases with missing data. It can pursue decision modeling while simultaneously imputing any missing clinical findings.

There are several promising future directions for this research project. First, it would be interesting to evaluate if the CSNN can actually improve the diagnostic performance of physicians by maintaining high sensitivity in predicting lesion malignancy while significantly reducing the number of unnecessary benign biopsies. Second, considering the recent trend to boost CAD performance by combining decision models, CSNN is a great candidate to join a pool of expert systems such as backpropagation neural networks, case based reasoning, and linear statistical techniques due to its inherently different theoretical foundation. A unified CAD tool that combines various decision models that "think" differently has a better chance to succeed since the models may complement each other. Finally, some exciting studies are possible if the CSNN is applied beyond diagnosis. Due to its inherent system optimization framework, the CSNN can be used with complete databases that include information on the patient's prognosis, treatment planning, and survival. For a given patient, the CAD operator can choose which neurons need prediction (i.e., diagnosis, treatment planning, prognosis). The CSNN can adapt easily to the decision task as dictated by the CAD operator because it has a non-hierarchical architecture. Ultimately, we envision this CAD tool developed on a large, comprehensive, and population diverse database, helping clinicians individualize the whole process of diagnostic management by exploring several hypothetical scenarios and choosing the one that optimizes survival outcome.

6. References

1 H.N. Schaller, "Constraint Satisfaction Problems," in *Optimization Algorithms*, edited by C.T. Leondes, (Academic Press, San Diego, CA, 1998), 209-248.

2 C.T. Chen, E.C.-K. Tsao, W.C. Tsao, "Medical image segmentation by a constraint satisfaction neural network," *IEEE Trans Nucl Sci* 38, 678-686, (1991).

3 A.J. Worth, D.N. Kennedy, "Segmentation of magnetic-resonance brain images using analog constraint satisfaction neural networks," *Image and Vision Computing* 12, 345-354, (1994).

4 F. Kurgollus, B. Sankur, "Image segmentation based on multi-scan constraint satisfaction neural network," *Pattern Recognition Letters* 20, 1553-1563, (1999).

5 American College of Radiology, "Breast Imaging Reporting and Data System," Reston, VA: American College of Radiology, (1996).

6 R.M. Golden, "Deterministic Nonlinear Dynamical Systems Analysis," in *Mathematical Methods for Neural Network Analysis and Design*, edited by R.M. Golden, (The MIT Press, Cambridge, MA, 1996), 115-142.

7 D.E. Rumelhart, P. Smolensky, J.L. McClelland, G.E. Hinton, "Schemata and sequential thought processes," in *Parallel Distributed Processing: Explorations in the Microstructures of Cognition (Vol. 2).* edited by D.E. Rumelhart and J.L. McClelland (The MIT Press, Cambridge, MA, 1986), 7-75.

8 M. Buscema, M. Intraligi, R. Bricolo, "Artificial neural networks for drug vulnerability recognition and dynamic scenarios simulation," *Substance Use & Misuse* 33, 587-623, (1998).

9 G. Massini, L. Shabtay, "Use of a constraint satisfaction network model for the evaluation of the methodone treatments of drug addicts," *Substance Use & Misuse* 33, 625-656, (1998).

10 G.D. Tourassi, M.K. Markey, J.Y. Lo, C.E. Floyd, Jr. "A Neural Network Approach to Breast Cancer Diagnosis as a Constraint Satisfaction Problem," *Medical Physics* 28(3):804-811, 2001.

11 C.E. Metz, B.A. Herman. J.H. Shen, "Maximum likelihood estimation of receiver operating characteristic curves from continuously-distributed data," Stat Med **17**, 1033-1053 (1998).

12 D.D. Adler and M.A. Helvie, "Mammographic biopsy recommendations," *Current Opinion in Radiology* 4, 123-129 (1992).

13 D.B. Kopans, "The positive predictive value of mammography," *AJR Am J Roentgenol* 158, 521-526 (1992).

14 J.Y. Lo, J.A. Baker, P.J. Kornguth, and C.E. Floyd, Jr, "Effect of patient history data on the prediction of breast cancer from mammographic findings with artificial neural networks," *Acad Radiol* 6, 10-15 (1999).

15 C.E. Floyd, Jr, J.Y. Lo, and G.D. Tourassi, "Cased-based reasoning computer algorithm that uses mammographic findings for breast biopsy decisions," *AJR Am J Roentgenol* 175, 1347-1352 (2000).

7. Appendices

A. **G.D. Tourassi**, M.K. Markey, and J.Y. Lo, "Validation of a constraint satisfaction neural network for breast cancer diagnosis: New results from 1030 cases," accepted at the 2003 SPIE Medical Imaging Conference, San Diego, CA, 15-20 February.

B. M.K Markey, J.Y. Lo, **G.D. Tourassi**, C.E. Floyd, Jr., "Self-Organizing Map for Cluster Analysis of a Breast Cancer Database," accepted for publication in *Artificial Intelligence in Medicine* [12/02].

Validation of a Constraint Satisfaction Neural Network for Breast Cancer Diagnosis: New Results From 1,030 Cases

Georgia D. Tourassi[1], Joseph Y. Lo[1], and Mia K. Markey[2]

[1]Department of Radiology, Duke University Medical Center, Durham, NC 27710
[2]Department of Biomedical Engineering, University of Texas, Austin, TX 78712

ABSTRACT

Previously, we presented a Constraint Satisfaction Neural Network (CSNN) to predict the outcome of breast biopsy using mammographic and clinical findings. Based on 500 cases, the study showed that CSNN was able to operate not only as a predictive but also as a knowledge discovery tool. The purpose of this study is to validate the CSNN on a database of additional 1,030 cases. An auto-associative backpropagation scheme was used to determine the CSNN constraints based on the initial 500 patients. Subsequently, the CSNN was applied to 1,030 new patients (358 patients with malignant and 672 with benign lesions) to predict breast lesion malignancy. For every test case, the CSNN reconstructed the diagnosis node given the network constraints and the external inputs to the network. The activation level achieved by the diagnosis node was used as the decision variable for ROC analysis. Overall, the CSNN continued to perform well over this large dataset with ROC area of Az=0.81±0.02. However, the diagnostic performance of the network was inferior in cases with missing clinical findings (Az=0.80±0.02) compared to those with complete findings (Az=0.84±0.03). The study also demonstrated the ability of the CSNN to effectively impute missing findings while performing as a predictive tool.

Keywords: computer-aided diagnosis, neural networks, breast cancer, constraint satisfaction

1. INTRODUCTION

Mammography is considered the most effective technique for early breast cancer diagnosis. Patients with early-detected malignancies have a significantly better prognosis [1,2]. Accordingly, physicians err on the side of caution and typically refer to biopsy all patients with unresolved suspicious findings in their diagnostic mammograms. However, the majority of biopsies (65-85%) performed due to suspicious mammograms are found to be benign [3-6]. The economic cost, physical burden, and emotional stress associated with excessive biopsy of benign lesions have been reported before [7-15]. Furthermore, another well-documented problem is the variability among radiologists regarding the clinical management (biopsy vs. follow-up) of suspicious breast lesions [16-19].

The application of computational techniques for the diagnostic interpretation of mammograms is one of the most active fields of research. The end product is typically a computer-aided decision (CAD) tool aimed to provide physicians with a reliable second opinion during their decision to biopsy a breast lesion. In a previous study, we developed a Constraint Satisfaction Neural Network (CSNN) to predict the outcome of breast biopsy based upon mammographic and clinical findings [20]. In a clinical setting, such predictive tool could assist radiologists in their decision to refer a patient suspected with breast cancer to biopsy or short-tern follow-up. Our studies showed that the CSNN allows us to explore predictive modeling as the optimization of a non-linear dynamic system [20]. Furthermore, the CSNN was used not only as a predictive tool but also as a flexible knowledge discovery tool decoding hidden data trends and associations. These studies were based on a limited set of 500 patients with complete mammographic and clinical findings. However, it remained uncertain whether the CSNN could be useful in larger patient samples with incomplete findings.

In the present study, we collected 1,030 consecutive clinical cases and used them as a validation test for the CSNN. First, we trained the CSNN on the original 500 cases. Then, we tested if the CSNN can achieve clinically acceptable diagnostic accuracy on the validation set. In addition, the effect of missing data was evaluated in more detail.

2. MATERIALS AND METHODS

2.1 The Constraint Satisfaction Network

The CSNN architecture has been described in detail before [20]. The CSNN is an auto-associative, Hopfield-type network [21] with neurons arranged in a non-hierarchical structure (Figure 1). Therefore, contrary to traditional predictive models, the CSNN does not have designated input and output neurons. The neurons are connected with symmetrical, bidirectional weights (w_{ij}=w_{ji}) but there are no reflexive weights (w_{ii}=0). The CSNN network operates as a non-linear, dynamic system aimed to achieve global stability by determining the activation status of its neurons while the weights remain fixed. The CSNN weights describe the problem constraints while every network state is a possible solution to the problem. A problem is solved when the network achieves a globally stable state without violating the constraints.

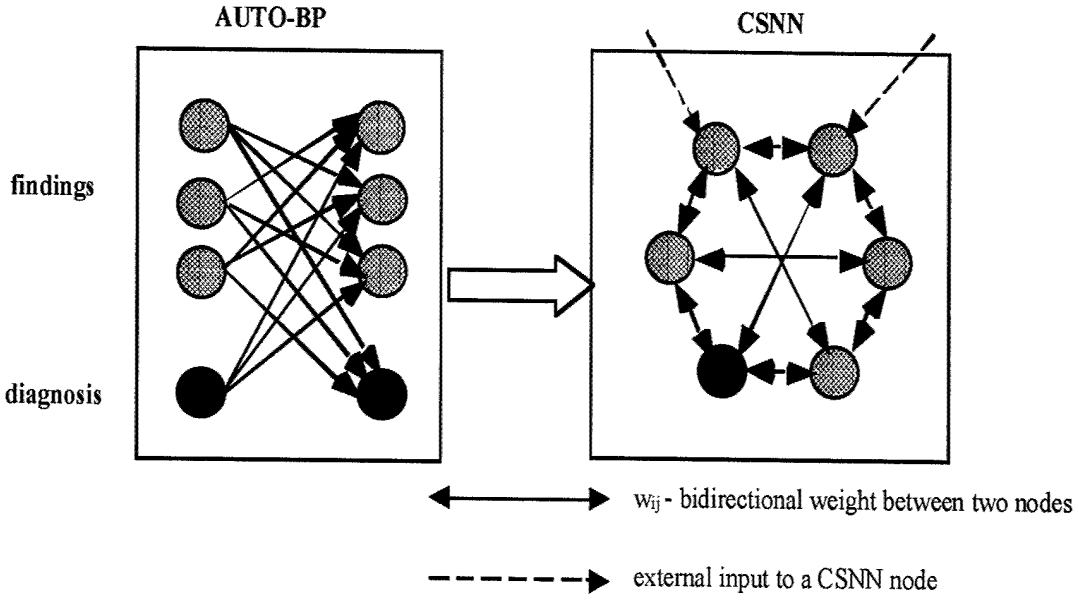

Figure 1: The CSNN architecture with the autoassociative backpropagation (auto-BP) training scheme

To achieve global stability, the CSNN employs a dynamic and iterative mechanism. The mechanism assumes that the activation level of all neurons can take any value in the range [0,1]. The CSNN is designed to maximize the activation of its neurons in relation to the constraints existing among them. To achieve this goal, the activation level of each neuron *i* is updated using the delta rule introduced by Rumelhart [22]. With this update rule, the network will restrict the activation levels to the [0,1] range and will evolve so that all neurons achieve their maximum possible activation while still satisfying the constraints imposed by the weights. The measure of global stability is a Lyaponov function often used to describe the state of nonlinear dynamic systems [21]. A dynamic system achieves a stable state when this function (known as Energy) is minimized. In the CSNN context, the energy function is a measure of constraint satisfaction.

A crucial step for developing a CSNN is determining the constraints weight matrix. The weight matrix contains the relations or constraints among all neurons. For this study we applied an autoassociative backpropagation (auto-

BP) scheme. The auto-BP network is a simple perceptron without hidden layers. The input and output layers have an equal number of nodes (N). During the training phase, the auto-BP learns to map any given pattern to itself using the backpropagation technique for gradient descent with the sigmoid activation function. When the training phase is complete, the autoassociative BP weights act as the CSNN constraints. Utilizing a backpropagation scheme to determine the CSNN constraints is highly innovative, overcoming the limitations of hard constraints typically associated with constraint satisfaction problems.

2.2 Data

The dataset consisted of non-palpable, mammographically suspicious breast lesions that underwent biopsy (core or excisional) at Duke University Medical Center from 1991 to 2000. There were in total 1,530 breast lesions with definitive histopathological diagnosis. The first 500 lesions (biopsied between 1991 and 1996) were used as the training set. The remaining 1,030 lesions (biopsied between 1996 and 2000) were consecutive cases and they were used as the validation set. The prevalence of breast cancer was the same (35%) in both sets. Table 1 provides some basic statistics regarding the training and validation sets. Breast lesions identified as "neither" in Table 1 represent special cases such as architectural distortion, regions of asymmetric breast density, areas of focal asymmetric density, and areas of asymmetric breast tissue.

Table 1: Comparison of the train and validation datasets

Data Set	Train Set	Validation Set
Total Number of Cases	500	1,030
Malignancies	174 (35%)	359 (35%)
Mean Age (yr)	55.5	55.9
Age Range	24-86	23-89
Mass cases	46%	39%
Calcification Cases	38%	47%
Masses with calcifications	6%	5%
Neither	9%	9%

Mammographic and clinical data were collected for each breast lesion according to collection procedures described before [20]. Briefly, for every lesion, expert mammographers reported the mammographic findings according to the BI-RADS lexicon [23]. Each BIRADS finding (with the exception of mass size) has a categorical rating. A higher rating typically represents a higher likelihood of malignancy. Patient age and history findings were also collected. In total, sixteen mammographic and clinical findings were recorded for each patient. Table 2 lists the findings selected to describe each case.

Complete mammographic and clinical findings were available for all 500 breast lesions in the train set. In the validation set, there were only 244 lesions (32.4% malignancy rate) with complete findings. For the remaining 786 lesions (35.5% malignancy rate), there were only mammographic findings available plus the patient's age at the time of diagnosis. The remaining clinical and history findings were unavailable for those patients.

All findings were converted into a binary input vector. A separate CSNN neuron was assigned to each possible rating for every finding. The two continuous findings (age and mass size) were represented as categorical data [20]. Specifically, mass size was coded in seven possible nodes. Each node corresponded in mass size increments of 10 mm. Similarly, patient age was coded in five nodes (<40yrs, 40-50, 50-60, 60-70, and >70 yrs old). In addition, one extra

neuron was added to constitute the diagnosis. The diagnosis neuron took the value of 1 if breast cancer was present and the value of 0 if breast cancer was absent. We used only a single diagnosis node so that the CSNN can be used as a predictive rather than a classification tool. In total, 83 CSNN neurons were used to represent the problem.

Table 2: Findings used to represent a breast lesion

Mammographic Findings	Value Range	Clinical Findings	Value Range
1. Calcifications Distribution	0-5	11. Patient Age	years
2. Calcifications Number	0-3	12. Family Hx of BC	0-1
3. Calcification Morphology	0-14	13. Personal Hx of BC	0-1
4. Quadrant Location of Abnormality	0-4	14. Hx of Benign Biopsy	0-1
5. Associated Findings	0-9	15. Menopausal Status	0-1
6. Special Cases	0-4	16. Hormone Therapy	0-1
7. Mass Margin	0-5		
8. Mass Shape	0-4		
9. Mass Density	0-4		
10. Mass Size	mm		

2.3 Performance Evaluation

During the development or "training" phase, the CSNN constraints were determined using the backpropagation autoassociative (auto-BP) network. The auto-BP network had an input layer and an output layer of 83 nodes each. Initially, the weights were randomly initialized and biases were set to 0. The auto-BP was then trained according to the backpropagation algorithm using the train set (i.e., the first 500 breast lesions). After the auto-BP weights and biases were determined, the weights served as the CSNN constraints. Next, the CSNN was applied as a predictive tool on the validation set (i.e., the remaining 1,030 lesions). For each test case, the CSNN network was used to predict the biopsy result based on the network's constraints (the weight matrix determined by auto-BP) and the external inputs (the available medical findings for each case). If a particular finding was present, then the corresponding external influence was active and set equal to 1.0. Initially, the activation levels of all CSNN neurons were randomly initialized. Then, available patient findings served as external inputs. The diagnosis neuron did not accept any external information and it was left to evolve based only on internal influences. Similarly, if there were missing clinical and history findings, then the corresponding neurons were left to evolve without any external influences.

At each iteration, the CSNN energy function was monitored to determine the stability of the network. In the end of the iterative process, the activation level achieved by the diagnosis neuron was used as the decision variable for Receiver Operating Characteristics (ROC) analysis. We used the ROCKIT software package developed by Metz *et al.* (http://xray.bsd.uchicago.edu/krl/toppage11.htm) to fit ROC curves to the activation level achieved by the CSNN diagnosis neuron.

3. RESULTS

3.1 Diagnostic Performance

The results of the validation study are summarized in Table 3. The table shows the overall ROC area index A_Z of the CSNN along with the partial area above a sensitivity of 90% ($_{0.90}A_Z$). The partial ROC area index for the high sensitivity range is a clinically more meaningful performance index for this diagnostic problem. The table also includes the CSNN positive predictive value (PPV) at 95% sensitivity. For comparison the Table also includes the previously reported CSNN performance on the initial 500 cases according to a 50%-50% cross-validation sampling scheme.

Table 3: Diagnostic Performance of the CSNN on the Initial and Validation Sets

Data Set	Az ±STD	$_{0.90}$Az±STD	PPV at 95% Sensitivity
Initial (500 cases)	0.84±0.02	0.35±0.06	50%
Validation (1,030 cases)	0.81±0.02	0.26±0.03	41%

In addition, the CSNN performance was analyzed separately according to the types of breast lesions (Table 4). Previous studies with a variety of artificial intelligence techniques have demonstrated that diagnostic performance varies substantially between masses and classifications [20,24,25]. Specifically, CAD performance on breast masses is superior to that on calcifications. Similar trend was observed in our validation study as well. CSNN performance was significantly better on masses than on calcifications. However, compared to the previous study [20], the CSNN performance deteriorated slightly on masses but improved on calcifications.

Table 4: CSNN diagnostic performance based on the type of lesions present.

Type of Lesions	No. of Cases (% maligancy) INITIAL SET	No. of Cases (% maligancy) VALIDATION SET	Az INITIAL SET	Az VALIDATION SET
Masses only	232 (29.7%)	402 (35.6%)	0.93±0.02	0.88±0.02
Calcifications only	192 (37.5%)	483 (31.5%)	0.65±0.04	0.70±0.03
Masses w/ Calcifications	29 (62.1%)	54 (50.0%)	0.83±0.08	0.75±0.07
No Masses or Calcifications	47 (31.9%)	91 (38.5%)	0.70±0.09	0.82±0.05

3.2 Effect of Missing data

As explained in the data description, the majority (786/1,030) of breast lesions in our validation database are missing the patients' clinical findings. The ROC area index of the CSNN was evaluated separately on the cases with complete findings and on those with incomplete clinical findings. As expected, the ROC area index was lower in

cases with missing data (A_Z=0.80±0.02) than in those with complete findings (A_Z=0.84±0.03). Similar trend was observed with the partial ROC area indices. Table 5 summarizes these findings.

Table 5: Effect of Missing Data on the Diagnostic Performance of the CSNN.

Cases	Number (% malignancy)	Az ± STD	$_{0.90}$Az ± STD	PPV at 95% Sensitivity
Complete	244 (32.4%)	0.84 ± 0.03	0.27 ± 0.08	41.6%
Incomplete	786 (35.5%)	0.80 ± 0.02	0.25 ± 0.04	39.5%

The next table presents the effect of missing data in more detail, according to the type of breast lesions present. The table shows that the presence of missing data reduces overall CSNN diagnostic performance. However, the difference was not statistically significant. A notable difference in performance was observed for the "masses with calcifications" category. However, the small number of cases in this category does not allow conclusive remarks. This is also the case for lesions without masses or calcifications present ("neither").

Table 6: CSNN performance according to the type of lesions present for cases with complete and incomplete findings

	Masses only	Calcifications only	Masses+Calcifications	Neither	ALL
Initial set	0.93 ± 0.02	0.65 ± 0.04	0.83 ± 0.08	0.70 ± 0.09	0.84 ± 0.02
Validation set	0.88 ± 0.02	0.70 ± 0.03	0.75 ± 0.07	0.82 ± 0.05	0.81 ± 0.02
Complete	0.91 ± 0.03	0.73 ± 0.07	1.0	0.81 ± 0.12	0.84 ± 0.03
Incomplete	0.87 ± 0.02	0.70 ± 0.03	0.63 ± 0.09	0.83 ± 0.05	0.80 ± 0.02

3.3 Ability to Impute Missing Data

The non-hierarchical architecture of the CSNN makes possible its utilization on cases with partially missing data. Other predictive models require an additional technique to impute the missing data before a case is tested. Contrary, the CSNN does not require such step. Specifically, the CSNN can be applied to reconstruct simultaneously not only the correct diagnosis but also any missing components of a given clinical case. This is an exciting possibility for clinical databases with missing data such as in our study. Imputing missing data is an important issue that tends to compromise the performance of a decision model.

We tested the accuracy of the CSNN to impute missing data while performing as a diagnostic tool. We focused on imputing the patient age. Previous studies have shown that the patient age is the strongest predictive clinical factor of malignancy [26]. We tested the CSNN on the same 1,030 validation cases. However, the CSNN neurons that represent patient age were left to evolve without any external influences. Therefore, we simulated an experiment where the CSNN was asked to perform as a diagnostic tool while imputing simultaneously a very important predictive finding (i.e., patient age).

Although the overall performance of the CSNN deteriorated (Az=0.78±0.02), it was still able to predict breast lesion malignancy with sufficient accuracy. Furthermore, the CSNN was able to impute the missing patient age

accurately in 30% of the cases. In 69% of the cases, the CSNN imputed patient age within adjacent age groups. Table 7 summarizes the results of this experiment. The table shows the true and CSNN predicted age groups for all patients in the validation set.

Table 7: CSNN accuracy on imputing the missing patient age while performing the diagnostic task

Patient Age Groups	No. of cases in each age group	Accuracy	Accuracy (± 1 age group)
<= 40 yrs	82	14.6%	51.2%
(40,50]	321	32.0%	69.5%
(50-60]	276	37.3%	68.8%
(60,70]	196	18.4%	85.7%
>70 yrs	155	34.8%	56.8%
TOTAL	1,030	29.9%	69.0%

4. DISCUSSION

In a previous study, we demonstrated the potential of using the Constraint Satisfaction Neural Network as a predictive and data mining tool for breast cancer diagnosis. The study utilized a cross-validation sampling scheme and a limited dataset of 500 breast lesions. The purpose of the present study was to validate the CSNN on a separate dataset of consecutive cases.

Overall, the CSNN performed well on the validation set as in the previous limited study. The previously reported trend of significantly better performance with masses than calcifications was successfully verified in the validation study. Some deterioration in performance was observed. However, the inferior performance can be attributed to two main factors. First, the validation set included more calcification than mass cases. Second, the majority of the validation cases had missing clinical findings. The effect of missing findings was studied in detail. The CSNN ability to effectively impute missing clinical data while performing as a predictive tool was verified successfully.

To summarize, the study reaffirmed the potential of using the CSNN as an effective predictive tool in breast cancer diagnosis. The ability to use the CSNN as predictive tool while simultaneously imputing any missing clinical findings makes the CSNN a promising alternative network for computer-aided diagnosis.

5. ACKNOWLEDGEMENTS

This work was supported by the U.S. Army Medical Research and Materiel Command grant DAMD17-01-1-0516.

6. REFERENCES

1. S. Shapiro, "Screening: assessment of current studies," *Cancer* 74, 231-238 (1994).
2. A. L. M. Verbeek, J. H. C. L. Hendriks, R. Holland, M. Mravunac, F. Sturmans, and N. E. Day, "Reduction of breast cancer mortality through mass screening with modern mammography," *Lancet* 1, 1222-1224 (1984).
3. D. D. Adler, and M. A. Helvie, "Mammographic biopsy recommendations," *Current Opinion in Radiology* 4, 123-129 (1992).
4. D. B. Kopans, "The positive predictive value of mammography," *AJR Am J Roentgenol* 158, 521-526 (1992).

5. S. Ciatto, L. Cataliotti, and V. Distante, "Nonpalpable lesions detected with mammography: review of 512 consecutive cases," *Radiology* 165, 99-102 (1987).
6. Knutzen AM, and Gisvold JJ, "Likelihood of malignant disease for various categories of mammographically detected, nonpalpable breast lesions," *Mayo Clin Proc* 68, 454-460 (1993).
7. Bassett LW, Bunnell DH, Cerny JA, and Gold RH, "Screening mammography: referral practices of Los Angeles physicians," *AJR Am J Roentgenol* 147, 689-692 (1986).
8. F. M. Hall, "Screening mammography - potential problems on the horizon," *NEJM* 314, 53-55 (1986).
9. F. M. Hall, J. M. Storella, D. Z. Silverstone, and G. Wyshak, "Nonpalpable breast lesions: recommendations for biopsy based on suspicion of carcinoma at mammography," *Radiology* 167, 353-358 (1988).
10. Cyrlak D, "Induced costs of low-cost screening mammography," *Radiology* 168, 661-3 (1988).
11. Sickles EA, "Periodic mammographic follow-up of probably benign lesions: results in 3,184 consecutive cases," *Radiology* 179, 463-468 (1991).
12. Varas X, Leborgne F, and Leborgne JH, "Nonpalpable, probably benign lesions: role of follow-up mammography," *Radiology* 184, 409-414 (1992).
13. Helvie MA, Ikeda DM, and Adler DD, "Localization and needle aspiration of breast lesions: complications in 370 cases," *AJR Am J Roentgenol* 157, 711–714 (1991).
14. Dixon JM and John TG, "Morbidity after breast biopsy for benign disease in a screened population," *Lancet* 1, 128 (1992).
15. Schwartz GF, Carter DL, Conant EF, Gannon FH, Finkel GC, and Feig SA, "Mammographically detected breast cancer: nonpalpable is not a synonym for inconsequential," *Cancer* 73,1660–1665 (1994).
16. Bird RE, Wallace TW, and Yankaskas BC, "Analysis of cancer missed at screening mammography," *Radiology* 184, 613-617 (1992).
17. Burhenne HJ, Burhenne LW, Goldberg D, Hislop TG, et al., "Interval breast cancer in screening mammography program in British Columbia: analysis and calcification," AJR *Am J Roentgenol* 162, 1067-1071 (1994).
18. Elmore J, Wells M, Carol M, Lee H, et al., "Variability in radiologists' interpretation of mammograms," *New England J Med* 331, 1493-1499 (1994).
19. Berg WA, Campassi C, Langenberg P, Sexton MJ, "Breast imaging reporting and data system: Inter- and intraobserver variability in feature analysis and final assessment,"AJR *Am J Roentgenol* 174, 1769-1777 (2000).
20. Tourassi GD, Markey MK, Lo JY, and C.E. Floyd, Jr. "A Neural Network Approach to Breast Cancer Diagnosis as a Constraint Satisfaction Problem," *Med Phys* 28, 804-811, (2001).
21. Golden RM, "Deterministic Nonlinear Dynamical Systems Analysis," in *Mathematical Methods for Neural Network Analysis and Design*, edited by R.M. Golden, (The MIT Press, Cambridge, MA, 1996), 115-142.
22. Rumelhart DE, Smolensky P, McClelland JL, Hinton GE, "Schemata and sequential thought processes," in *Parallel Distributed Processing: Explorations in the Microstructures of Cognition (Vol. 2).* edited by D.E. Rumelhart and J.L. McClelland (The MIT Press, Cambridge, MA, 1986), 7-75.
23. American College of Radiology, "Breast Imaging Reporting and Data System," Reston, VA: American College of Radiology, (1996).
24. Bilska-Wolak A, Floyd CE Jr., "Breast Biopsy predictions using a case-based reasoning classifier for masses versus classifications," *SPIE Proceedings*, Medical Imaging 2002, Vol. 4684, 661-665 (2002).
25. Markey MK, Lo JY, Floyd CE, "Differences between computer-aided diagnosis of breast masses and that of calcifications," *Radiol* 223: 489-493 (2002).
26. Lo JY, Baker JA, Kornguth PJ, Floyd CE, "Effect of patient history data on the prediction of breast cancer from mammographic findings with artificial neural networks," *Acad Radiol* 6, 10-15 (1999).

Self-Organizing Map for Cluster Analysis of a Breast Cancer Database

Mia K. Markey[1,2], Joseph Y. Lo[1,2], Georgia D. Tourassi[2], and Carey E. Floyd Jr.[1,2]

[1] Department of Biomedical Engineering, Duke University, Durham, North Carolina 27708

[2] Digital Imaging Research Division, Department of Radiology, Duke University Medical Center, Durham, North Carolina 27710

Corresponding author and reprint address:

Mia K. Markey
ENS617B C0800
Department of Biomedical Engineering
The University of Texas at Austin
Phone: +1.512.471.1711
Fax: +1.512.471.0616
email: mia.markey@mail.utexas.edu

Abstract

The purpose of this study was to identify and characterize clusters in a heterogeneous breast cancer computer-aided diagnosis database. Identification of subgroups within the database could help elucidate clinical trends and facilitate future model building. A Self-Organizing Map (SOM) was used to identify clusters in a large (2258 cases), heterogeneous computer-aided diagnosis database based on mammographic findings (BI-RADS™) and patient age. The resulting clusters were then characterized by their prototypes determined using a Constraint Satisfaction Neural Network (CSNN). The clusters showed logical separation of clinical subtypes such as architectural distortions, masses, and calcifications. Moreover, the broad categories of masses and calcifications were stratified into several clusters (seven for masses, three for calcifications). The percent of the cases that were malignant was notably different among the clusters (ranging from 6% to 83%). A feed-forward back-propagation artificial neural network (BP-ANN) was used to identify likely benign lesions that may be candidates for follow up rather than biopsy. The performance of the BP-ANN varied considerably across the clusters identified by the SOM. In particular, a cluster (#6) of mass cases (6% malignant) was identified that accounted for 79% of the recommendations for follow up that would have been made by the BP-ANN. A classification rule based on the profile of cluster #6 performed comparably to the BP-ANN, providing approximately 25% specificity at 98% sensitivity. This performance was demonstrated to generalize to a large (2177) set of cases held-out for model validation.

Index terms: self-organizing map, cluster analysis, breast cancer, computer-aided diagnosis

1. Introduction

There is considerable interest in the use of computational techniques to aid in the detection and diagnosis of breast cancer [5, 8, 26]. Most computer-aided diagnosis (CAD) studies, including this one, focus on mammography since it is the primary tool for the detection of breast lesions and the subsequent decision to biopsy suspicious lesions. The decision to biopsy is complicated by the fact that breast cancer can present itself in a variety of ways on a mammogram and there is considerable overlap in the appearance of benign and malignant lesions. CAD systems for the decision to biopsy that are based on findings extracted by radiologists are often trained and evaluated over heterogeneous databases that reflect this variability in the morphological appearance of suspicious breast lesions [1, 7, 28]. We have recently shown that a CAD tool trained on such a heterogeneous database can perform very differently on two broad subgroups which constitute most of the currently biopsied lesions: masses and microcalcifications [17]. In particular, we observed that the performance was significantly better on masses than on calcifications.

In this study, we used a self-organizing map (SOM) [13] to identify clusters in a heterogeneous breast cancer CAD database. SOM is an unsupervised learning method that relates similar input vectors to the same region of a map of neurons. To the best of our knowledge, SOMs have not been used to identify clusters in a CAD database similar to the one presented here. SOMs have been used for other tasks in breast cancer CAD such as a benchmark for model selection [27] and to predict biopsy outcome [4].

Once the SOM was used to identify the clusters, a constraint-satisfaction neural network (CSNN) was used to characterize the clusters by determining a profile for each

cluster. Briefly, the CSNN is a Hopfield-type network of neurons arranged in a non-hierarchical way (Figure 1). There are symmetric, bidirectional weights between all pairs of neurons but there are no reflexive weights. The CSNN operates as a nonlinear, dynamic system that tries to reach a globally stable state by adjusting the activation levels of the neurons under the constraints imposed by the *a priori* fixed weight values. A cluster "profile" provides a description of a "typical" case in the cluster. We have previously introduced CSNN for predicting biopsy outcome and as a data mining tool for breast cancer CAD databases [25].

A feed-forward back-propagation artificial neural network (BP-ANN) is a classic technique that is commonly used in breast cancer CAD systems. Consequently, a BP-ANN was used to predict the biopsy outcome [2, 10, 21] and the performance of the BP-ANN was compared on the clusters identified by the SOM and profiled by the CSNN.

A clustering algorithm such as an SOM followed by a cluster characterization method such as CSNN profiling could serve as tools in the initial phases of a divide-and-conquer approach to the computer-aided diagnosis of breast cancer. Both modular and ensemble methods could be used for a divide-and-conquer approach. A modular system uses multiple classifiers to solve a classification problem by partitioning the input space into smaller domains, each of which is handled by a local model [24]. The local models can be thought of as experts for a particular kind of case. Ensemble methods are resampling schemes in which the same cases are used in training multiple experts, whose predictions are then combined [24]. Such approaches may be justified in light of recent results in this field. Simple ensembles of classifiers using voting or averaging to combine their predictions have shown promise in computer-aided detection of breast masses [14,

22, 31]. Zheng *et al.* employed a modular scheme, in which the data were partitioned by a difficulty measure, for computer-aided detection of breast masses with encouraging results [30]. Zheng *et al.* also investigated a promising ensemble of modular models, formed by taking the average of the predictions from modular models in which the data were partitioned using three features [29]. Huo *et al.* described a modular system, in which the data were partitioned by a spiculation measure, which was superior to a general image-based computer-aided diagnosis system [11, 12]. Finally, we have recently demonstrated that a BI-RADS™-based CAD tool built on a heterogeneous database can perform very differently on two broad subgroups of lesions, masses and microcalcifications [17]; the CAD tools investigated performed better on masses than on calcifications. In all of the examples listed here, *a priori* knowledge was used to partition the data into subsets. Unsupervised learning may provide an alternate avenue to *a priori* knowledge for identifying subsets in the data that should be handled separately in the development or evaluation of computer-aided diagnosis or detection systems.

2. Materials and Methods

2.1. Data

Approximately half of the available data (4435) were used for model development (2258) in this study in order to withhold the remaining data for additional model validation (2177); the data were randomly partitioned into the training and validation sets, but attention was paid to key summary statistics such as the fraction of cases that were malignant in each set. For each lesion, the benign or malignant status from pathologic diagnosis was known. The overall malignancy fraction was 43%. In the next

few paragraphs, we describe the data (2258) used for model development in greater detail.

The first data set consisted of 751 non-palpable, mammographically suspicious breast lesions that underwent biopsy (core or excisional) at Duke University Medical Center from 1990 to 2000. The data collection procedures have been previously described [16]. Briefly, expert mammographers described each case using the Breast Imaging and Reporting Data System (BI-RADS™) lexicon [20]. Each of the cases was read by one of 7 readers. When a lesion could be described by multiple descriptors (e.g., pleomorphic and punctate), the mammographers were requested to report the descriptor that was most suspicious for malignancy (e.g., pleomorphic). Of the 751 cases, 260 (35%) were malignant.

The second data set consisted of 501 mammographically suspicious breast lesions that underwent excisional biopsy at the University of Pennsylvania Medical Center from 1990 to 1997. The data collection procedures have been previously described [16]. Briefly, each of the cases was read by one of 11 expert mammographers who described each case using the BI-RADS™ lexicon [20]. When a lesion could be described by multiple descriptors (e.g., pleomorphic and punctate), the mammographers were requested to report the descriptor that was most suspicious for malignancy (e.g., pleomorphic). Of the 501 cases, 200 (40%) were malignant.

The third data set consisted of 1006 biopsy-proven breast lesions randomly selected from the Digital Database for Screening Mammography [9]. Expert mammographers described each case using the BI-RADS™ lexicon [20]. Lesions that were described by multiple descriptors were encoded for our purposes using the

descriptor that was most suspicious for malignancy. Of the 1006 cases, 522 (52%) were malignant.

Specifically, the six BI-RADS™ features collected describe the mass margin, mass shape, calcification morphology, calcification distribution, associated, and special findings. Missing values were encoded as zero. Each BI-RADS™ feature was encoded using uniformly scaled rank ordered categories (Table 1). For example, when a mass is present for a case, the mass margin can take on one of five values: well circumscribed (1), microlobulated (2), obscured (3), ill-defined (4), or spiculated (5). In addition to the BI-RADS™ features, the patient age was collected, for a total of seven features.

2.2. Self-Organizing Map

A self-organizing map relates similar cases (input vectors) to the same region of a map of neurons [13]. The SOM was computed using the SOM toolbox in MATLAB® (The MathWorks Inc., Natick, MA). The basic SOM consisted of 16 neurons arranged in a single layer in a 2-D square grid of 4 by 4 neurons, but different configurations were considered. For each case, the Euclidean distance between the case and each neuron was calculated based on the seven input features (the biopsy outcome was not provided to the SOM). For input to the SOM, each feature was scaled by subtracting the mean and dividing by the standard deviation, resulting in each scaled feature having mean zero and standard deviation of one. After the most similar neuron was determined the neurons in its neighborhood were identified. The neighborhood of a neuron was defined as all the neurons within a given link distance of the matched neuron. All the neurons in the neighborhood were adjusted to have feature values closer to the current case. The amount that the neuron weights were adjusted was controlled by the learning rate. The

learning rates and distance threshold values used were the default values for the SOM toolbox.

2.3. Constraint Satisfaction Neural Network

After the clusters were identified, a CSNN was used to determine the profiles of the clusters [23, 25]. Custom software in the C language was used to implement the CSNN and has been previously described [25]. The Lyapunov energy function was used as a measure of the network stability. It was found that 1000 iterations were sufficient to achieve stability. The weights were predetermined using autoassociative backpropagation neural networks (auto-BP). In keeping with our previous work [25], the auto-BP networks were trained with a learning rate of 1.0 for 100 iterations and the root mean squared training error was approximately 0.1 (network outputs between 0 and 1).

For each cluster, a CSNN was used to generate a profile. Each category of the categorical BI-RADS™ features corresponded to a binary variable and associated neuron. For example, the mass margin with its five non-zero categories was represented by five separate neurons. Patient age was translated into a discrete variable with five levels (< 40 years, $40 \leq x < 50$, $50 \leq x < 60$, $60 \leq x < 70$, ≥ 70 years) [25]. An additional neuron was used to signify cluster membership. The activation level of the neuron indicating cluster membership was set to the maximal value and the other neurons were allowed to evolve until the network reached a stable state. The feature neurons that were activated defined the profile of the cluster. A profile is a list of feature values that succinctly summarizes the cluster and defines a "typical" case (e.g., mass margin is well circumscribed, mass shape is round, and patient age is between 50 and 60 years). All cases in the cluster don't exactly match the profile; there is still a distribution of feature values. Notice that unlike

common summary statistics, such as the cluster centroid, the CSNN profile implicitly includes feature selection; only features deemed relevant to the network for describing a cluster are included.

2.4. Back-Propagation Artificial Neural Network (BP-ANN)

A feed-forward back-propagation artificial neural network (BP-ANN) was used to predict the biopsy outcome from the mammographic findings and patient age. The BP-ANN was trained to minimize the sum-of-squares error using the back-propagation algorithm [2, 10, 21]. The network had a single hidden layer of 14 neurons and each neuron in the network used a logistic activation function. The network inputs (7) were the BI-RADS™ features and patient age. Network inputs were rescaled to 0 to 1 (by subtracting the minimum value and dividing by the maximum minus the minimum). The biopsy outcomes were the network targets; there was one output node indicating malignancy. The 2258 cases were presented to the network in a round-robin manner (leave-one-out, k-fold cross-validation with k = N) and training ended before the average testing error on the left-out cases began to increase. The custom neural network software used was written in C++ by members of our laboratory, and the training and testing process has been reported previously [15, 17].

2.5. Receiver Operating Characteristic

Receiver Operating Characteristic (ROC) curves can be used to show the trade-off in sensitivity and specificity achievable by a classifier by varying the threshold on the output decision variable [18, 19]. The area under the ROC curve is often used as a measure of classifier performance. In evaluating models for diagnosing breast cancer, all sensitivities are not of equal interest. Only techniques that perform with very high

sensitivity would be clinically acceptable since missing a cancer (false negative) is generally considered much worse that an unnecessary benign biopsy (false positive). Thus, particular attention was paid to the specificity at 98% sensitivity.

The ROC curves were calculated non-parametrically. P-values and standard deviations on the specificity at 98% sensitivity were estimated by bootstrap sampling on the decision variable [6].

3. Results

Figure 2 illustrates the arrangement of the neurons in the SOM. The set of cases that were mapped to a neuron defined a cluster. Figure 2 shows the number of cases that were mapped to each neuron, i.e., the number of cases in each cluster. The fraction of the cases in each cluster that were malignant is also shown in Figure 2 (bottom number in italics). The malignancy fraction is not shown for the clusters with fewer than 10 cases (#5, 12, and 15), on the assumption that no meaningful conclusions can be drawn from such a small number of cases. Inspection of the cases mapped to these clusters (#5, 12, and 15) revealed that the cases are rare for this database. They included cases with findings that were seen with a very low prevalence in the set (*e.g.*, special finding of intramammary lymph node) or reflected incomplete or inconsistent data (*e.g.*, the calcification morphology was described but calcification distribution feature was not reported). Together these three clusters comprise only 0.5% of the cases. Therefore, no further analysis was performed on these clusters. Recall that the SOM was not provided with the biopsy outcome information. The differences in the malignancy fraction are a reflection of differences in the BI-RADS™ features and patient age between the clusters. Cluster malignancy rates near 50% do contain some information since the overall

malignancy fraction was 43%. Notice that there is generally a higher incidence of malignant lesions in the clusters on the righthand side of the map.

Figure 3 shows the effect that changing the SOM architecture has on the clusters identified. Alternative architectures allow one to vary the number of neurons as well as their topological layout, thus potentially allowing for variations in the complexity of the model. One alternative to a 4 x 4 SOM is a smaller but still square 3 x 3 SOM (Figure 3a). In figure 3b, the clusters of the 3 x 3 and 4 x 4 SOMs are compared using a bubble plot. For each case, the neuron it mapped to was determined for each SOM. The number of cases for each pair of clusters between the two SOMs was plotted; the size of the circle indicates the number of cases. The more large bubbles that are present in such a plot, the more the SOMs agreed on the clustering of the cases. Similarly, figures 3c and 3d show the comparison with a 5 x 5 SOM. Linear trends (i.e., bubbles lining up along the diagonals) indicate that the same cases are being mapped to the same region (e.g., upper right-hand area) in the two SOMs. In addition to square topologies, other layouts were also investigated which utilized approximately the same number of neurons. Comparisons were made to a 2 x 8 SOM and to a three-dimensional SOM of 2 x 3 x 3 neurons, both with approximately the same number of neurons as the 4 x 4 square SOM.

For the 4 x 4 SOM, the cluster profiles generated by the CSNN are shown in Figure 4. Each cell in the table represents the feature categories that were dominant or most strongly associated with the cases matching that cluster. Profiles were not computed for the clusters with very few cases. The mass cases are distributed over neurons #2, 3, 4, 6, 7, and 8. The profiles of neurons #9, 13, 14, and 16 indicate that those clusters contain microcalcifications. Neuron #1's profile indicates that that cluster is comprised of focal

asymmetric densities. Note that the profile for neuron #10 includes only the age variable. The profile for neuron #11 reveals that the lesions in that cluster are architectural distortions.

An alternative approach to generating cluster profiles is to compute summary statistics such as the feature mode (or mean for real-valued features such as age). Figure 5 shows the mode profiles of the clusters identified by the 4 x 4 SOM. For the most part, there is considerable agreement between the CSNN and mode profiles. Most of the differences correspond to adjacent categories in the features (Table 1) where the CSNN has selected the second most prevalent value for the profile. However, using multiple methods to summarize the clusters may be beneficial. For example, the CSNN profile of neuron #16 (Figure 4) does not include any mass features yet the feature mode profile (Figure 5) shows that the mass features are usually non-zero. In fact, inspection of the cases in the cluster defined by neuron #16 reveals that they are calcified masses. Conversely, the CSNN profile for neuron #10 (Figure 4) includes only the age variable while the mode profile's (Figure 5) inclusion of values for the calcification variables may be misleading for this small cluster (N = 29) where there is little dominance by any single value.

A BP-ANN was trained to predict the biopsy outcome from the BI-RADS™ features and patient age. Figure 6 shows the ROC curve for the BP-ANN. The SOM can also be used to generate a malignancy prediction [4]. For each case, the prediction was the fraction of the cases that were malignant in the cluster that the case was mapped to by the SOM. For example, if a case belonged to cluster #4 in which 83% of the cases were malignant, then the classifier output for that case would be 0.83. Notice that using this

approach limits the number of operating points on the non-parametric ROC curve to the number of clusters with unique malignancy fractions minus one (Figure 6). The performance at the highest sensitivities was comparable. In particular, at 98% sensitivity the SOM operates with 0.26 ± 0.03 specificity and the BP-ANN operates with 0.25 ± 0.03 specificity ($p = 0.93$).

Figure 7 lists how the BP-ANN trained on all the cases performs in terms of the BP-ANN's recommendations for follow up instead of biopsy on the subsets identified by the SOM. A threshold was applied to the BP-ANN outputs such that the overall sensitivity was approximately 98% (965/982) with resulting specificity of approximately 24% (303/1276). In other words, 320 cases (303 actual negatives and 17 actual positives) fell below the threshold. These 320 cases that the BP-ANN would have recommended for follow up are shown in Figure 7 according to which SOM cluster they belonged. Notice that there is considerable variability in the performance on the clusters. In particular, the majority of the cancers that the BP-ANN would have referred to follow up (11/17 = 65%) and the majority of the benign lesions that the BP-ANN would have spared biopsy (242/303 = 80%) were in the cluster defined by neuron #6.

These interesting results with the cluster defined by neuron #6 suggested that a simple rule-based approach could be valuable. We developed a classification rule based on the cluster profiles (Figures 4 and 5) of neuron #6 and a Classification And Regression Tree (CART) [3] model for mass cases using the implementation in S-PLUS® (Insightful Corp., Seattle, WA). The classification rule was: if the mass margin was well-circumscribed or obscured and the age was less than 59 years and there were no calcifications, associated findings, or special findings, then don't biopsy, otherwise do

biopsy. On the 2258 training cases, this rule gave 961 / 982 = 98% sensitivity and 336 / 1276 = 26% specificity. In other words, this rule performed comparably to the BP-ANN with a threshold of 0.1842 (965 / 982 = 98% sensitivity, 303 / 1276 = 24% specificity).

The performance of the BP-ANN and the classification rule developed from data mining were evaluated on the 2177 cases withheld for model validation. On the validation set, the classification rule gave 886 / 904 = 98% sensitivity and 339 / 1273 = 27% specificity and the BP-ANN with a threshold of 0.1842 gave 884 / 904 = 98% sensitivity and 296 / 1273 = 23% specificity. Thus, both the BP-ANN and the rule-based approach generalized and they performed comparably at this high sensitivity point.

4. Discussion

Considerable variability was seen in the fraction of the cases that were malignant from cluster to cluster. Several clusters had malignancy fractions that were notably different from the fraction of the entire data set (43%). One of the major goals of computer-aided diagnosis of breast cancer is to identify very likely benign cases as candidates for follow-up in lieu of biopsy, in order to reduce the number of benign biopsies. Therefore, the clusters with very low malignancy fractions (e.g., neuron #6 with 6% malignant) are dominated by such very likely benign lesions and may be of particular interest for further studies. It is possible to use the clusters and their malignancy fractions directly as a tool for predicting biopsy outcome [4]. For each case, the prediction was the fraction of the cases that were malignant in the cluster that the case was mapped to by the SOM (Figure 6). For very high sensitivities, this prediction scheme (98% sensitivity, 0.26 ± 0.03 specificity) was competitive with the back-propagation artificial neural network (98% sensitivity, 0.25 ± 0.03 specificity, p = 0.93);

however, this SOM-based method was not superior to the BP-ANN. The SOM prediction method in conjunction with the CSNN profiling method has the potential advantage that physicians may understand the intuition behind it better than they do the BP-ANN, which is often seen as a "black box". The SOM prediction method, similar to a case-based reasoning system, predicts the probability of malignancy of a new case by reporting the fraction of similar cases that were found to be malignant [7]. The SOM prediction method could also potentially be used in an ensemble of classifiers. If the outputs of two classifiers are not strongly correlated, it is possible that they could be combined to produce a classifier that is better than either of its component classifiers.

The effects of the changing the SOM architecture were investigated (Figure 3). As indicated by the presence of large circles in the bubble plots, the SOMs with similar architectures showed substantial agreement in clustering the data. Moreover, the presence of linear trends in the comparisons with the 5 x 5, 2 x 8, and 2 x 3 x 3 SOMs suggest that similar SOM architectures result in similar geometric relationships between clusters. These data argue that the clustering is relatively insensitive to the SOM architecture for this problem. It should be noted that this study did not focus on the organization of the clusters into a topological map. Consequently, many of the analyses in this study could have been performed using other clustering algorithms.

Figure 4 lists the CSNN profiles for the clusters identified with the SOM. The successful separation of *a priori* known, coarse lesion types (masses, clustered microcalcifications, focal asymmetric densities, and architectural distortions) provided some quality assurance of the clustering. Clusters were further identified within the general group of mass lesions, reflecting different combinations of the mass margin, mass

previous work demonstrating performance differences with an *a priori* partitioning of the data into two broad subgroups of lesions, masses and microcalcifications [17] and suggests that further work should be done to investigate building cluster-specific models. The variation in the BP-ANN performance across the clusters could also influence the ultimate clinical implementation of the decision aid since it may not be useful to apply the BP-ANN to cases similar to those groups of cases for which it always recommended biopsy in the training set. Interestingly, the SOM identified a cluster of mass cases (#6) which accounted for the majority cases that the BP-ANN would have recommended for follow up rather than biopsy. Recall that the identification of likely benign cases that could be spared biopsy is the goal of such computer-aided diagnosis schemes. This suggests that the SOM clustering and CSNN profiling technique could be used to provide the physician with an alternative description of what the BP-ANN does for certain types of cases. The identification of a single cluster that accounted for the majority of the cases that the BP-ANN would have recommended for follow up also suggests the investigation of rule-based methods to identify relatively simple diagnostic criteria which might be applied to these cases to aid the radiologists in their decision making process. Based on the profiles of the clusters identified by the SOM, we developed a simple classification rule that performed comparably to the BP-ANN (approximately 25% specificity with 98% sensitivity). Moreover, we demonstrated that the classification rule generalized to 2177 cases withheld for model validation.

5. Acknowledgements

This work was supported in part by Susan G. Komen Breast Cancer Foundation grant DISS0100400, U.S. Army Medical Research and Materiel Command grants DAMD17-

shape, and patient age variables. The cluster profiles that included calcification features showed stratification of the general group of calcification lesions only by patient age and not any of the calcification findings. Notice that while some features may not be considered useful by the CSNN for profiling individual clusters, it is possible that they could be useful to other summarizing techniques or to methods designed to describe the differences between clusters.

An alternative approach to characterizing the clusters is to calculate summary statistics for each of the features. Figure 5 shows the mode for each of the BI-RADS™ features and the mean of the patient age for each cluster. In general, there is good agreement in the cluster descriptions obtained from these summary plots and the CSNN profiles. However, they are not identical. The most notable differences are for neurons #10 and #16, which show the advantages and disadvantages respectively of the fact that the CSNN method inherently includes feature selection.

It may be easier to interpret a CSNN profile, with typically only a few dominant features per cluster, than to interpret as many summary values as there are input findings. Note as well that the CSNN takes into the account interdependencies between the features, while the summary statistics were based on each feature independently. CSNN profiles or summary statistics can be used to quickly sort through the results of a clustering technique, but additional characterization may be appropriate for clusters of particular interest.

Classification based on the SOM was competitive to that achieved by the BP-ANN at high sensitivity levels (Figure 6). Notable variation in the performance over the clusters identified by the SOM was observed (Figure 7). This is consistent with our

02-1-0373 and DAMD17-01-1-0516, and NIH/NCI R29 CA-75547. We would like to thank Brian Harrawood for scientific programming.

6. References

[1] J. A. Baker, P. J. Kornguth, J. Y. Lo, M. E. Williford and C. E. Floyd, Jr., Breast cancer: prediction with artificial neural network based on BI-RADS standardized lexicon, Radiology 196 (1995) 817-22.
[2] C. M. Bishop, Neural Networks for Pattern Recognition (Oxford University Press, New York, 1995).
[3] L. Breiman, J. Friedman, R. Olshen and C. Stone, Classification and Regression Trees (Wadsworth International Group, Belmont, 1984).
[4] D. Chen, R. F. Chang and Y. L. Huang, Breast cancer diagnosis using self-organizing map for sonography, Ultrasound in Medicine & Biology 26 (2000) 405-11.
[5] K. Doi, H. MacMahon, S. Katsuragawa, R. M. Nishikawa and Y. Jiang, Computer-aided diagnosis in radiology: potential and pitfalls, European Journal of Radiology. 31 (1999) 97-109.
[6] B. Efron and R. J. Tibshirani, An Introduction to the Bootstrap (Chapman & Hall, New York, NY, 1993).
[7] C. E. Floyd, Jr., J. Y. Lo and G. D. Tourassi, Case-based reasoning computer algorithm that uses mammographic findings for breast biopsy decisions, American Journal of Roentgenology 175 (2000) 1347-52.
[8] M. L. Giger, Computer-aided diagnosis of breast lesions in medical images, Computing in Science & Engineering 2 (2000) 39-45.
[9] M. Heath, K. W. Bowyer and D. Kopans, Current status of the Digital Database for Screening Mammography, in: N. Karssemeijer, M. Thijssen and J. H. Hendriks, eds., Digital Mammography (Kluwer Academic Publishers, Dordrecht, 1998) 457-460.
[10] J. Hertz, K. Anders and R. G. Palmer, Introduction to the Theory of Neural Computation (Addison-Wesley, Redwood City, California, 1991).
[11] Z. Huo, M. L. Giger and C. E. Metz, Effect of dominant features on neural network performance in the classification of mammographic lesions, Physics in Medicine & Biology. 44 (1999) 2579-95.
[12] Z. Huo, M. L. Giger, C. J. Vyborny, D. E. Wolverton and C. E. Metz, Computerized classification of benign and malignant masses on digitized mammograms: a study of robustness, Academic Radiology 7 (2000) 1077-84.
[13] T. Kohonen, Self-Organizing Maps (Springer-Verlag, Berlin, 1995).
[14] L. Li, Y. Zheng, L. Zheng and R. A. Clark, False-positive reduction in CAD mass detection using a comptetive classification strategy, Medical Physics 28 (2001) 250-258.
[15] J. Y. Lo, J. A. Baker, P. J. Kornguth and C. E. Floyd, Jr., Effect of patient history data on the prediction of breast cancer from mammographic findings with artificial neural networks, Academic Radiology. 6 (1999) 10-5.
[16] J. Y. Lo, M. K. Markey, J. A. Baker and C. E. Floyd, Jr., Cross-institutional evaluation of BI-RADS predictive model for mammographic diagnosis of breast cancer, American Journal of Roentgenology 178 (2002) 457-63.

[17] M. K. Markey, J. Y. Lo and C. E. Floyd, Jr., Differences between computer-aided diagnosis of breast masses and that of calcifications, Radiology 223 (2002) 489-93.
[18] C. E. Metz, Basic principles of ROC analysis, Seminars in Nuclear Medicine 8 (1978) 283-98.
[19] C. E. Metz, ROC methodology in radiologic imaging, Investigative Radiology. 21 (1986) 720-33.
[20] Illustrated breast imaging reporting and data system (BI-RADS™) (American College of Radiology, Reston, VA, 1998).
[21] D. E. Rumelhart and J. L. McClelland, Parallel Distributed Processing: Explorations in the Microstructures of Cognition Volume 1: Foundations (The MIT Press, Cambridge, Massachusetts, 1986).
[22] R. Rymon, B. Zheng, Y. H. Chang and D. Gur, Incorporation of a set enumeration trees-based classifier into a hybrid computer-assisted diagnosis scheme for mass detection, Acad Radiol 5 (1998) 181-7.
[23] H. N. Schaller, Constraint Satisfaction Problems, in: C. T. Leondes, ed., Optimization Techniques (Academic Press, San Diego, CA, 1998) 209-248.
[24] A. J. C. Sharkey, Combining Artificial Neural Nets: Ensemble and Modular Multi-Net Systems (Springer-Verlag, London, 1999).
[25] G. D. Tourassi, M. K. Markey, J. Y. Lo and C. E. Floyd, Jr., A neural network approach to breast cancer diagnosis as a constraint satisfaction problem, Medical Physics 28 (2001) 804-11.
[26] C. J. Vyborny, M. L. Giger and R. M. Nishikawa, Computer-aided detection and diagnosis of breast cancer, Radiologic Clinics of North America 38 (2000) 725-40.
[27] D. West and V. West, Model selection for a medical diagnostic decision support system: a breast cancer detection case, Artificial Intelligence in Medicine 20 (2000) 183-204.
[28] Y. Wu, M. L. Giger, K. Doi, C. J. Vyborny, R. A. Schmidt and C. E. Metz, Artificial neural networks in mammography: application to decision making in the diagnosis of breast cancer, Radiology 187 (1993) 81-7.
[29] B. Zheng, Y. Chang, W. F. Good and D. Gur, Performance gain computer-assisted detection schemes by averaging scores generated from artificial neural networks with adaptive filtering, Medical Physics 28 (2001) 2302-2308.
[30] B. Zheng, Y. H. Chang and D. Gur, Adaptive computer-aided diagnosis scheme of digitized mammograms, Academic Radiology 3 (1996) 806-14.
[31] B. Zheng, Y. H. Chang and D. Gur, Mass detection in digitized mammograms using two independent computer-assisted diagnosis schemes, American Journal of Roentgenology 167 (1996) 1421-4.

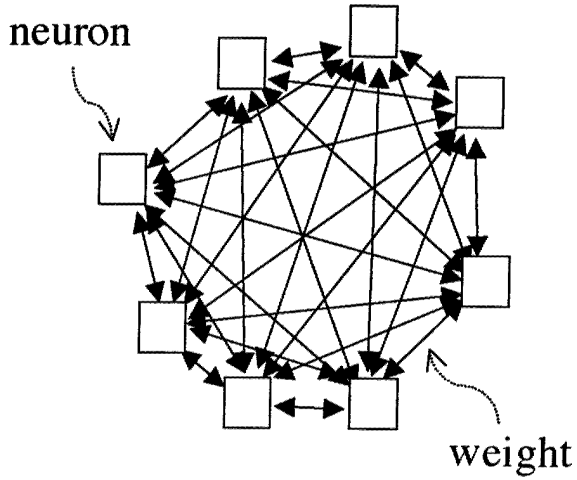

Figure 1. Schematic of the constraint satisfaction neural network (CSNN). Notice that the neurons are fully interconnected with no reflexive weights.

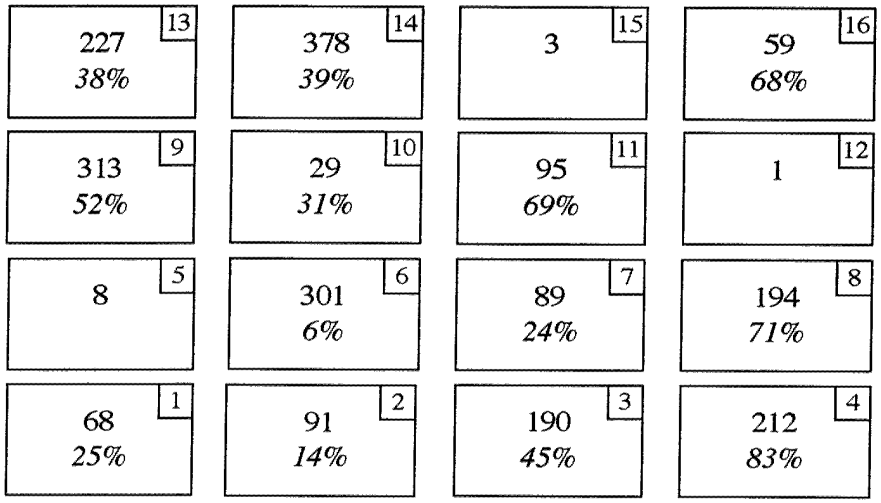

Figure 2. Index of the neurons in the 4 x 4 map. Each neuron defined a cluster. The number of cases that were mapped to each neuron, i.e., the number of cases in each cluster (normal type), and the fraction of the cases in each cluster that were malignant (*italics*) is shown. Malignancy fraction data not shown for the clusters with very few cases. Over all, 43% of the cases were malignant. Information regarding the main features of the cases in each cluster is shown in Figures 4 and 5.

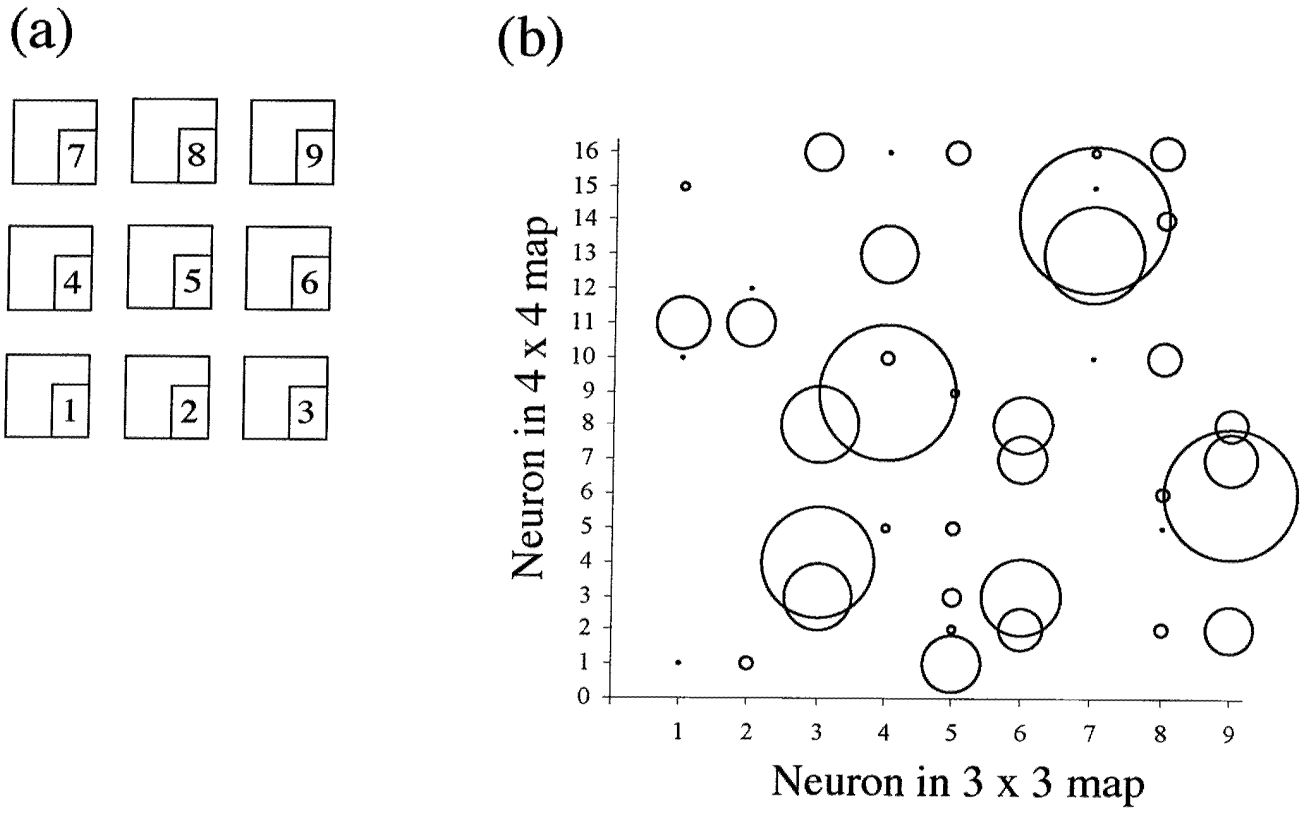

Figure 3. (a) The index of the neurons in the 3 x 3 map. (b) Comparison of the clusters identified by the 3 x 3 and 4 x 4 SOMs. For each case, the neuron it mapped to was determined for each SOM. The number of cases for each pair of clusters between the two SOMs was plotted; the size of the circle indicates the number of cases. The more large bubbles that are present in such a plot, the more the SOMs agreed on the clustering of the cases. Linear trends (i.e., bubbles lining up along the diagonals) indicate that the same cases are being mapped to the same region in the two SOMs.

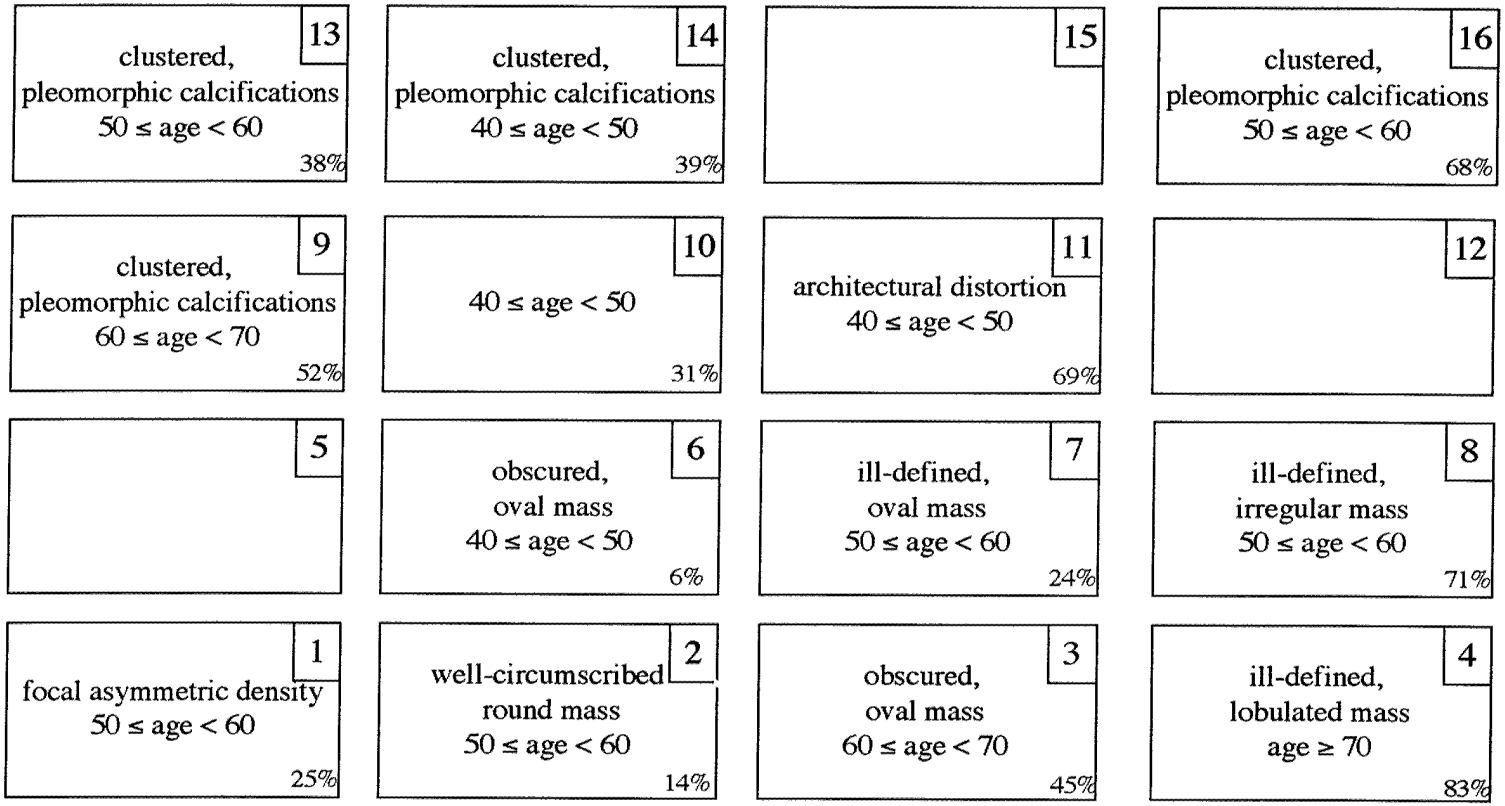

Figure 4. The cluster profiles generated by the CSNN for the clusters identified by the 4 x 4 SOM (cluster number in upper right corner). A cluster "profile" provides a description of a "typical" case in the cluster. Profiles were not computed for neurons #5, 12, and 15 which had very few cases mapped to them. The percent of the cases that were malignant is shown in the lower righthand corner; refer to Figure 2.

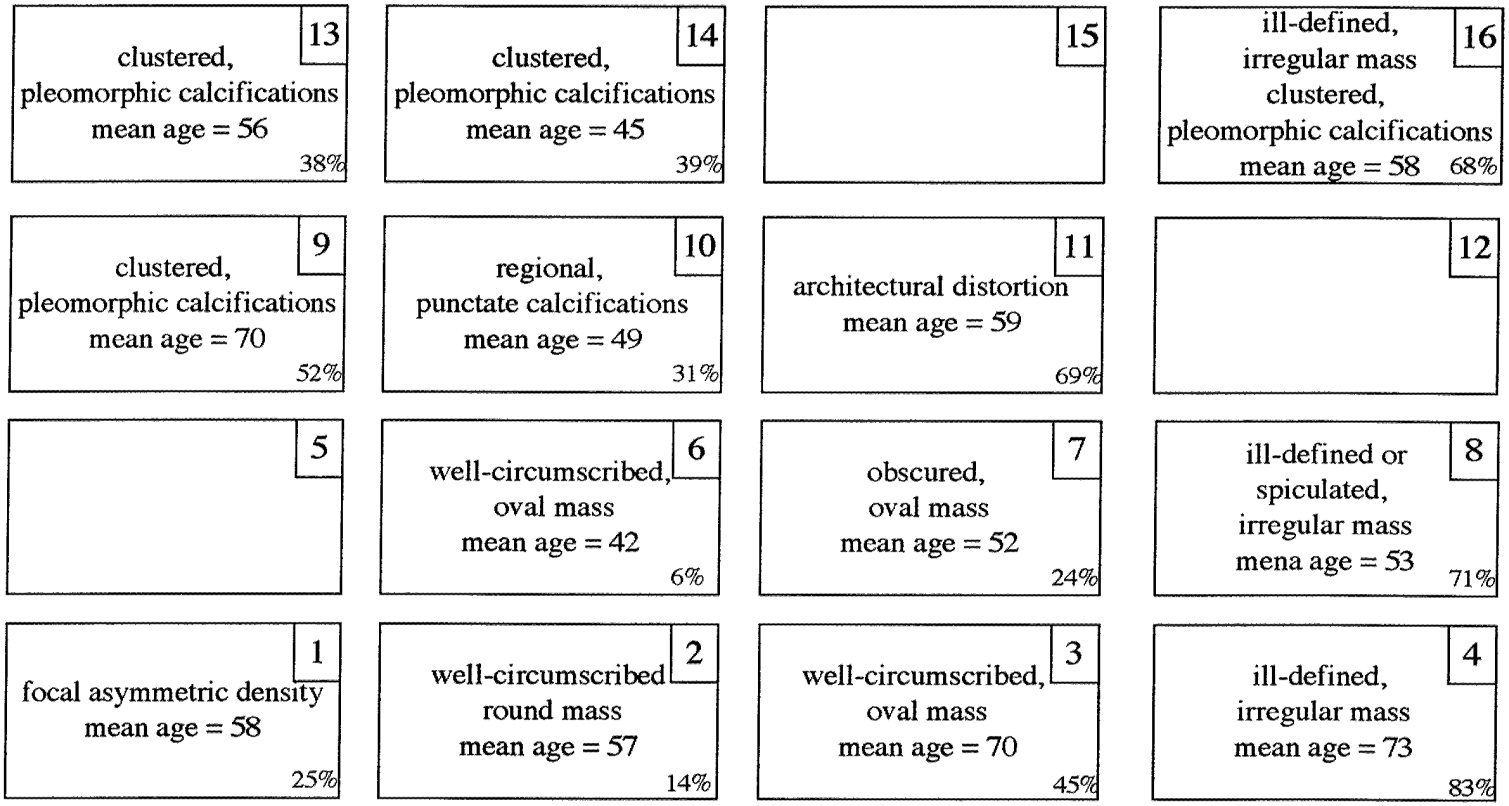

Figure 5. The cluster profiles generated by the computing the mode the features (mean for age) for the clusters identified by the 4 x 4 SOM (cluster number in upper right corner). A cluster "profile" provides a description of a "typical" case in the cluster. Profiles were not computed for neurons #5, 12, and 15 which had very few cases mapped to them. Features for which the mode value indicated that the feature was absent were omitted (e.g., mass margin = no mass). The percent of the cases that were malignant is shown in the lower righthand corner; refer to Figure 2.

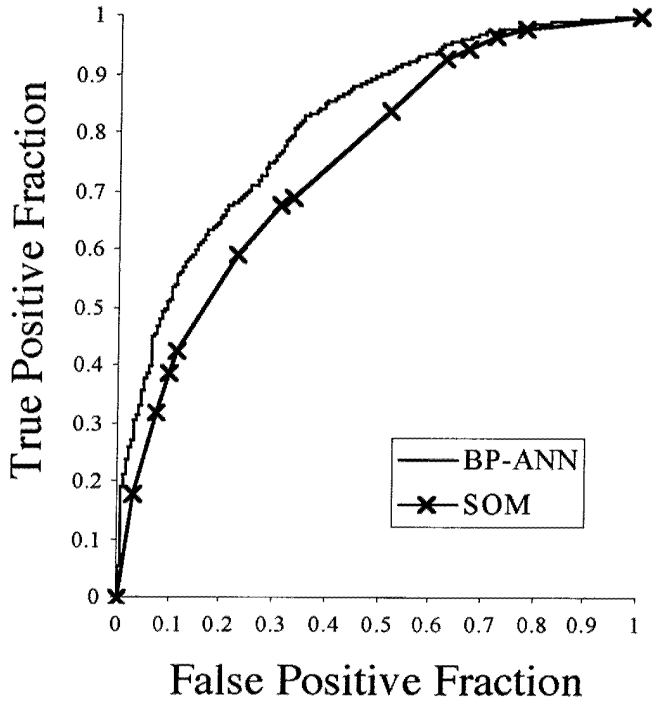

Figure 6. ROC curves for the BP-ANN and the SOM. For each case, the prediction from the SOM was the fraction of the cases in the cluster it belonged to that were malignant.

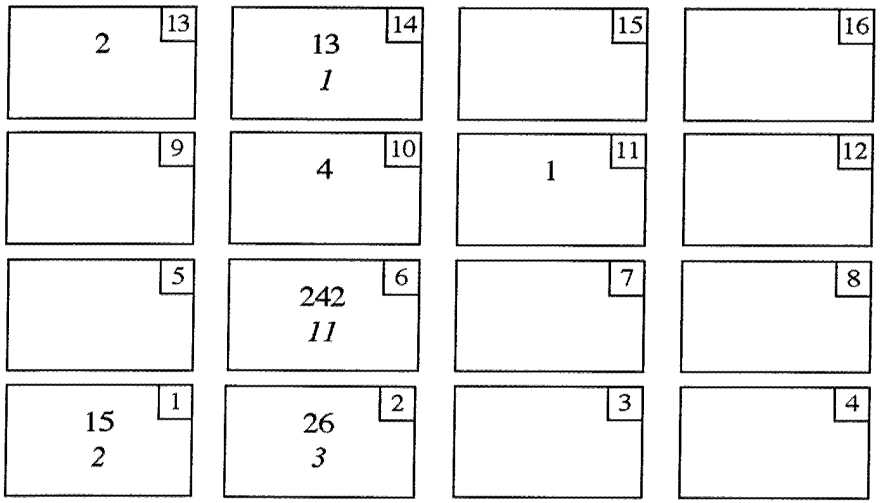

Figure 7. Comparison of the performance of the BP-ANN trained on all the cases on the clusters identified by the SOM. For each cluster the number of true negatives (normal type) and the number of false negatives (*italics*) is shown.

Table 1. Encoding of the BI-RADS™ features.

Mass margin	**Mass shape**	**Calcification morphology**	**Calcification distribution**	**Associated findings**	**Special findings**
0 – no mass	0 – no mass	0 – no calcifications	0 – no calcifications	0 - none	0 – none
1 – well circumscribed	1 - round	1 – milk of calcium like	1 - diffuse	1 – skin lesion	1- intrmam. lymph node
2 – mircolobolated	2 - oval	2 - eggshell or rim	2 - regional	2 - hematoma	2 – asymmetric breast tissue
3 - obscured	3 - lobulated	3 - skin	3 - segmental	3 – post surgical scar	3 – focal asymm. density
4 – ill-defined	4 - irregular	4 - vascular	4 - linear	4 - trabecular thickening	4 – tubular density
5 - spiculated		5 - spherical or lucent centered	5 - clustered	5 – skin thickening	
		6 - suture		6 – skin retraction	
		7 - coarse		7 – nipple retraction	
		8 – large rod-like		8 – axillary adenopathy	
		9 - round		9 – architectural distortion	
		10 - dystrophic			
		11 - punctate			
		12 - indistinct			
		13 - pleomorphic			
		14 – fine branching			